

DOCUMENT RESUME

ED 443 336

HE 033 092

AUTHOR Obiekwe, Jerry C.
TITLE The Multidimensional Character of Teaching Effectiveness: A Comparative Analysis of Student Evaluation Responses of Full and Part-Time Faculty.
PUB DATE 1999-10-00
NOTE 20p.; Paper presented at the Annual Conference of the Mid-Western Educational Research Association (Chicago, IL, October 13-16, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150) -- Tests/Questionnaires (160)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; College Students; Evaluation Methods; *Full Time Faculty; Higher Education; Multidimensional Scaling; *Part Time Faculty; Predictor Variables; Student Attitudes; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; *Test Reliability; *Test Validity
IDENTIFIERS *Multidimensionality (Tests); Unidimensionality (Tests)

ABSTRACT

This study compared college students' responses on their evaluations of the effectiveness of full- and part-time college faculty. A group of 1,101 students completed evaluation instruments for all courses taught by full-time faculty, and 2,067 students completed evaluations for all courses taught by part-time faculty in spring 1998. In fall 1998, 1,231 students completed the questionnaire for classes taught by full-time faculty, and 2,580 students completed it for classes taught by part-time faculty. The evaluation instrument consisted of 14 items on teacher effectiveness (e.g., clearly stated course objectives, genuine concern with student progress, well-prepared for class, encourage student questions, and accomplish course objectives). Students' responses were analyzed to investigate the multidimensionality, validity, and reliability of the instrument. Results indicate that the instrument was unidimensional with both full- and part-time faculty. The instrument's reliability, which was relatively high, was the same when used by full-time and part-time faculty to evaluate teaching effectiveness. Results suggest that the instrument could be used to predict student academic performance. The results also indicate that class size and student ratings are inversely related. Nine data tables are included. The questionnaire is appended. (Contains 27 references.) (SM)

**THE MULTIDIMENSIONAL CHARACTER OF TEACHING
EFFECTIVENESS: A COMPARATIVE ANALYSIS OF STUDENT
EVALUATION RESPONSES OF FULL AND PART-TIME FACULTY.**

**Jerry C. Obiekwe
The University of Akron-Wayne College**

PERMISSION TO REPRODUCE AND
TRANSMIT THIS MATERIAL HAS
BEEN GRANTED BY

J. OBIKWE

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

**Paper Presented at the Annual Conference of the Mid-Western Educational
Research Association. October 13-16, 1999, Chicago, Illinois.**

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

BEST COPY AVAILABLE

THE MULTIDIMENSIONAL CHARACTER OF TEACHING EFFECTIVENESS: A COMPARATIVE ANALYSIS OF STUDENT EVALUATION RESPONSES OF FULL AND PART-TIME FACULTY.

Introduction

It is now almost a commonplace practice in American universities and colleges for each course taught by instructional personnel to be evaluated by students. In Seldon's (1984) survey of American universities and colleges, he estimates that about 70% of institutions of higher learning collect data on student course evaluation. Also, in their survey of 40 research universities, Ory and Parker (1989) discovered that 100% of these institutions do collect data on student rating of quality teaching. These data are used in part in the determination of teaching excellence (Abrami, D'Apollonia, & Cohen, 1990; Marsh, 1987; Perry, 1990). Other uses of student evaluation, as cited by Marsh (1984), are as follows: (a) formative feedback to faculty about their teaching effectiveness; (b) decisions on tenure and promotion; (c) students' use in course selection and instructors; and (d) for research purposes.

Teaching excellence or quality of teaching is well established in the literature as a multidimensional construct (Abrami & D'Apollonia, 1990; Bolton, Bonae & Hinman, 1976; Fernandez & Mateo, 1992; Frey, 1978; Isaacson, Mckeachie, Milholland, Lin, Hofeller, Baerwaldt, & Zinn, 1964; Jeffreys, Massoni, O'Donnell, & Smoldaka, 1997; Theall & Franklin, 1990). For example, Fernandez and Mateo (1990) were able to extract two reliable factors which explained 71% of the total variance in their analysis of the student ratings of quality instruction. Because of the types of variables that loaded on the first factor, it was named teaching competence demonstrated, while the second factor was named motivational skills.

Abrami and D'Apollonia (1990) are of the opinion that six factors are usually generated from the students' evaluation data which describe teaching effectiveness. They called these factors skills, rapport, structure, difficulty, interaction, and feedback. The skills factor relates to the knowledge of the subject, while rapport describes the instructor's harmony with the students. The structure factor deals with the organization of the course, and the difficulty factor relates to the difficulty of the assignments. The interaction factor describes the instructor's skills in facilitating class discussions. The

feedback factor relates to how well the teacher keeps the students abreast of their progress.

In another study, Jeffreys et al. (1997) analyzed data collected from nursing students with three different student evaluation instruments. They were able to establish the multidimensional quality of teaching and learning with each of these instruments. Essentially, data collected with each instrument yielded more than one factor. Isaacson et al. (1964) extracted six factors in their student evaluation analysis which they labeled skill, overload, structure, feedback, group interaction, and student-teacher rapport.

There are other studies that have also given credence to the multidimensional concept of quality teaching (Marsh, 1991; Marsh & Hocevar, 1991; Marsh, 1984; Marsh, 1983). In each of these studies, the researchers consistently extracted nine factors in their analysis of student responses to Student Evaluation of Education Quality (SEEQ). El-Hassan (1995) obtained eight dimensions of teaching in her logical analysis of student responses to the Teaching Effectiveness Scale administered to the students of the American University of Beirut.

Clearly, the implication from the review of the literature relating to student evaluation is that teaching effectiveness has multidimensional characteristics. More importantly, these dimensions of teaching effectiveness have been shown in meta-analysis of multisection validity studies to be related to student academic achievement (Cohen 1981, 1983, 1986, 1987). However, according to Cohen's report, certain dimensions such as structure, interaction, skill, learning, rapport and evaluation may be more related to achievement than others. Other researchers have established that enthusiasm as a dimension of teaching effectiveness is highly correlated with student ratings of quality instruction (Doyle & Whitely, 1974).

Objective of the Study

The purpose of this paper is to present results of a comparative analysis between full-time and part-time faculty with regard to student evaluation responses. The comparison focused on the following research questions:

- (1) Are the dimensions of teaching effectiveness, as measured by the student evaluation instrument, invariant regardless of student evaluation data generated by students taught by full-time or part-time faculty?

- (2) Is the reliability of the instrument consistent regardless of student evaluation data generated by students taught by full-time or part-time faculty?
- (3) Is the predictive validity of the instrument consistent regardless of student evaluation data generated by students taught by full-time or part-time faculty?
- (4) Is the linear relation between student ratings and class size consistent regardless of student evaluation data generated by students taught by full-time or part-time faculty?

Instrument

The evaluation instrument consists of 14 items (see Appendix A) that were selected from a 35-item instrument previously used to evaluate teaching effectiveness at the college. The rationale behind reducing the number of items to 14 was to reduce the amount of time students use in completing the questionnaire. The question now is: Is the current instrument multidimensional? Is it valid and is it reliable?

For items 1-6 on the instrument, the student rates the teaching faculty on a five-point scale. The first possible answer on the scale is "strongly agree," which is rated 5. The second is "agree," rated 4; the third is "neither agree nor disagree," rated 3; the fourth is "disagree," rated 2; and the fifth possible answer is "strongly disagree," rated 1. For items 7-13, the five possible answers on the rating scale are worded differently; however, the scale is still based on five points. The following are the possible answers: "almost always" (rated 5), "often" (rated 4), "occasionally" (rated 3), "seldom" (rated 2), and "almost never" (rated 1). Item 14, which is the only global item in the instrument, has five possible answers that are worded differently but are still based on a five-point scale. The following are the five possible answers: "superior" (rated 5), "very good" (rated 4), "good" (rated 3), "fair" (rated 2), and "poor" (rated 1).

Subjects

There were 1101 students who completed the student evaluation instrument for all courses taught by full-time faculty and 2067 students for

all courses taught by part-time faculty in the spring semester of 1998. In the Fall 1998, there were 1231 students who completed the student questionnaire for classes taught by full-time faculty as opposed to 2580 students for courses taught by part-time faculty. Their responses were analyzed in order to investigate the multidimensionality, the validity, and the reliability of the instrument via the statistical package for social science (SPSS Version 8.0).

Procedure

In both spring and fall semesters of 1998, full-time faculty student evaluation data were separately analyzed from the part-time faculty data. The results of these analyses were then compared.

Dimension of the instrument

In order to examine the dimensionality of the instrument, the item mean ratings of each instructor were submitted to principal component analysis. Only one factor was extracted in each instance, and this factor was labeled "Teaching Competence." The extraction criteria were based on eigenvalue greater than or equal to one and the scree plot (Gorsuch, 1983). Table 1 displays the results of the factor pattern between Full-time and part-time faculty in both Spring and Fall 1998. In Table 1(Spring 1998), the total variance explained by the competence factor of the full-time faculty was 80%, while for part-time faculty it was 77%. Still in Table 1 (Fall 1998), the total variance explained by the competence factor for the full-time faculty was 86% compared to 78% for part-time faculty.

Reliability

The reliability of a student evaluation instrument is the ability of the items and the scales to produce consistent responses at all times in all courses. These consistent responses are as a result of the degree of agreement among respondents. The reliability of the instrument will be evaluated by the internal consistency of the 14 items which gives the lower bound of the actual reliability (Novick & Lewis, 1967). Nunnally (1978) defines internal consistency of an instrument as the average correlation among the items. This average correlation estimates the internal consistency of the instrument and is called coefficient alpha or Cronbach alpha. The coefficient alpha is an index that is between 0 and 1. Nunnally (1978) sets

the minimum acceptable index at .70. Since all the items loaded on one factor, showing the unidimensional nature of the instrument, its reliability can be established by taking a look at the internal consistency of the 14 items. Actual student evaluation responses were used in the calculation of the coefficient alpha. Table 2 shows the internal consistency coefficient of the evaluation instrument obtained by the separate use of full and part-time student evaluation response data in spring and fall 1998.

Table 1 . Comparative Factor Solutions for Full-time and Part-time Faculty.

Items	Competence Factor Full-time: Spring 1998	Competence Factor Part-time: Spring 1998	Competence Factor Full-time: Fall 1998	Competence Factor Part-time: Fall 1998
1	.92	.87	.94	.92
2	.75	.86	.90	.88
3	.85	.76	.93	.80
4	.86	.85	.83	.83
5	.91	.89	.98	.91
6	.93	.91	.93	.89
7	.78	.77	.87	.77
8	.95	.93	.96	.92
9	.95	.91	.94	.91
10	.92	.88	.94	.89
11	.83	.84	.90	.79
12	.95	.92	.96	.91
13	.94	.90	.95	.94
14	.96	.94	.98	.97

Table 2. Comparative Coefficient Alpha for Full-time and Part-time Faculty.

Full-time Faculty: Spring 1998	Part-time Faculty: Spring 1998	Full-time Faculty: Fall 1998	Part-time Faculty: Fall 1998
.95	.95	.96	.97

Predictive validity of the instrument

Predictive validation deals with the question: Can student ratings be used to predict criterion measures that are related to teaching effectiveness? In order to empirically test this question, Marsh (1984) stated the following:

(a) there should be many multisection courses available; (b) students should be randomly assigned to those course sections; (c) students should be pretested, and the scores should be correlated to the criterion measure; (d) the multisection courses should have a common syllabi with a common objective and a common final examination; (e) each section should be taught by a different teacher; (f) the common final exam should be designed by someone who did not teach the course. With these conditions, the predictive validity of the student ratings will be upheld if there exists a strong positive correlation between the section mean ratings and the criterion measure (final exam).

Since this investigation involves the analysis of pre-existing data, it simply examined the following correlations based on the fact that the dimensions of teaching effectiveness correlate with student academic performance (Marsh, 1984): (a) the Pearson correlation between the section mean ratings and the section mean achievement of courses whose number of sections are greater than or equal to three for Spring and Fall 1998 (see Table 3 and Table 4 respectively) and (b) the Pearson correlation between the section global mean ratings and section mean achievement of courses whose number of sections are greater than or equal to three for Spring and Fall 1998 (see Table 5 and Table 6 respectively). The 14th item in the instrument is the global item (see appendix A). Achievement in this study is defined as the final letter grade earned by students. Table 7 shows the Pearson correlation between the section mean ratings and the section achievement for selected courses, while Table 8 shows the Pearson correlation between the section mean ratings and the section achievement means for all courses.

Table 3. Pearson Correlation Between Section Mean Ratings and Section Mean Achievement of Courses Whose Number of Section is $n \geq 3$ for Fall 1998.

Course Name	Pearson Correlation Full-time: Fall 1998	Pearson Correlation Part-time: Fall 1998
All Sections	.25** n = 86	.18* n = 192
Software Fundamentals	.92* n = 4	.98 n = 3
English Composition I	.39 n = 6	.37 n = 11
College Algebra	-.07 n = 3	
Intro. To Psychology	-.84 n = 3	
Intro. To Sociology	.44 n = 3	
Effective Oral Communication	.80* n = 6	.93 n = 3
Human Relations		.85 n = 4
Intro. To Windows		-.73 n = 5
Spreadsheet Software		.94 n = 3
English Composition II		.88 n = 4
Geology		.29 n = 3
Preparatory Math.		.82** n = 8
Basic Writing II		-.45 n = 5
Basic Math I		-.32 n = 3
Basic Math II		.62 n = 5
Reading & Study Skills		-.83 n = 4
Fitness & Wellness		.00 n = 3

* $p < .01$ ** $p < .05$

Table 4. Pearson Correlation Between Global Mean Ratings and Section Mean Achievement of Courses Whose Number of Section is $n \geq 3$ for Fall 1998

Course Name	Pearson Correlation Full-time: Fall 1998	Pearson Correlation Part-time: Fall 1998
All Sections	.56** n = 86	.18* n = 192
Software Fundamentals	.91* n = 4	.97 n = 3
English Composition I	.19 n = 6	.53* n = 11
College Algebra	-.17 n = 3	
Intro. To Psychology	-.68 n = 3	
Intro. To Sociology	.20 n = 3	
Effective Oral Communication	.71 n = 6	.82 n = 3
Human Relations		.73 n = 4
Intro. To Windows		-.22 n = 5
Spreadsheet Software		.86 n = 3
English Composition II		.88 n = 4
Geology		.80 n = 3
Preparatory Math.		.76 n = 8
Basic Writing II		-.52 n = 5
Basic Math I		-.99 n = 3
Basic Math II		.56 n = 5
Reading & Study Skills		-.15 n = 4
Fitness & Wellness		.09 n = 3

* $p < .01$, ** $p < .05$

Table 5. Pearson Correlation Between Section Mean Ratings and Section Mean Achievement of Courses Whose Number of Section is $n \geq 3$ for Spring 1998.

Course Name	Pearson Correlation and number of sections Full-time: Spring 1998	Pearson Correlation and number of sections Part-time: Spring 1998
All Sections	.29* n = 83	
English Composition I	.92* n = 3	.44 n = 4
English Composition II	.96* n = 6	-.06 n = 5
Effective Oral Communication	0.00 n = 5	.68* n = 3
Software Fundamentals		.71* n = 5
Introduction to Windows		1.00* n = 3
Preparatory Mathematics		-.002 n = 5
Introduction to Sociology		.77* n = 3

* $p < .01$

Table 6. Pearson Correlation Between Global Mean Ratings and Section Mean Achievement of Courses Whose Number of Section is $n \geq 3$ for Spring 1998.

Course Name	Pearson Correlation and number of sections Full-time: Spring 1998	Pearson Correlation and number of sections Part-time: Spring 1998
All Sections	.18 n = 83	
English Composition I	.61* n = 3	.09 n = 4
English Composition II	.95* n = 6	.09 n = 5
Effective Oral Communication	-.33 n = 5	.68* n = 3
Software Fundamentals		.84* n = 5
Introduction to Windows		1.00* n = 3
Preparatory Mathematics		-.02 n = 5
Introduction to Sociology		.91* n = 3

* $p < .01$

Table 7. Pearson Correlation Between Section Mean Ratings and Section Achievement Means for Selected Courses.

Full-time Faculty: Spring 1998 English Composition II	Part-time Faculty: Spring 1998 Computer & Software Fundamentals	Full-time Faculty: Fall 1998 Effective Oral Communication	Part-time Faculty: Fall 1998 Preparatory Mathematics
.95* N = 6 P<.01	.71 N = 5 P<.01	.80* N = 6 P<.05	.82* N = 8 P<.01

Table 8. Pearson Correlation Between Section Mean Ratings and Section Achievement Mean for all Course Sections.

Full-time Faculty: Spring 1998 All Course Sections	Part-time Faculty: Spring 1998 All Course Sections	Full-time Faculty: Fall 1998 All Course Sections	Part-time Faculty: Fall 1998 All Course Sections
.20* N = 83 P<.05	.34* N = 144 P<.01	.25* N = 86 P<.01	.18* N = 192 P<.01

Student ratings and class size

In order to investigate the relationship between class size and student ratings, section mean ratings were correlated with the number of students in each section. In all instances, student ratings correlate negatively with class size (see Table 9).

Table 9. Pearson Correlation Between Student Ratings and Class Size.

Full-time Faculty: Spring 1998 All Course Sections	Part-time Faculty: Spring 1998 All Course Sections	Full-time Faculty: Fall 1998 All Course Sections	Part-time Faculty: Fall 1998 All Course Sections
-.23* N = 83 P<.05	-.12 N = 144 P<.01	-.18 N = 86 P<.01	-.27* N = 192 P<.01

Discussion and Conclusions

The first research question was: Are the dimensions of teaching effectiveness as measured by the student evaluation instrument invariant regardless of student evaluation data generated by students taught by full-time or part-time faculty? The result clearly demonstrated that the

instrument is unidimensional with full-time faculty and unidimensional with part-time faculty.

The second research question was: Is the reliability of the instrument consistent regardless of student evaluation data generated by students taught by full-time or part-time faculty? The reliability of the instrument, which was relatively high, appears to be same when used by the full-time faculty to evaluate teaching effectiveness or when used by the part-time faculty for the same purpose. Essentially, the 14 items and the scales will produce consistent responses at all times in all courses regardless of who is being evaluated.

The third research question was: Is the predictive validity of the instrument consistent regardless of student evaluation data generated by students taught by full-time or part-time faculty? The Pearson correlation between section mean ratings and section achievement means for selected courses was high and also significant. The correlation between the section mean ratings and the achievement mean for all course sections was low but significant; it was also relatively the same for students taught by full-time faculty and students taught by part-time faculty in the two semesters examined. With these consistent results, the implication here could well be that the instrument can be used to predict student academic performance.

The fourth research question was: Is the linear relation between student ratings and class size consistent regardless of student evaluation data generated by students taught by full-time or part-time faculty? The results in all instances, clearly showed that class size and student ratings seem to be inversely related.

The dimension of the instrument, the predictive validity, the reliability, and the correlation between class size and the student ratings were consistently replicated by the separate analyses of student evaluation data generated by students taught by full-time or part-time faculty in the spring and fall of 1998. Although the instrument may not be multidimensional, but the only extracted dimension, which was teaching competence, appears to be predictive of student academic performance. In summary, the reliability and the validity of the instrument have been upheld.

As a result of the outcomes from this study, the college's student evaluation committee has chosen to define teaching effectiveness as a

multidimensional construct and, consequently, wants to develop an instrument whose items would elicit multiple dimensions of quality of teaching. The committee plans on recommending to the faculty a type of customized evaluation instrument where each faculty member will be evaluated on three core dimensions and, perhaps, two other dimensions of their choice. The three core dimensions are believed to transcend all disciplines, while the other two are believed to be dimensions that are inherently specific to the faculty member's discipline. Alternatively, the committee could recommend that each faculty member be evaluated on five dimensions of his or her choice.

The proposed instrument will not only be used to measure teaching effectiveness, but will also be used to give faculty diagnostic feedback on their teaching. Since teaching is currently being redefined as a scholarly activity by the Carnegie Foundation for the Advancement of Teaching, coupled with increased attention on student assessment, finding ways to effectively evaluate and continuously improve our teaching is a critical issue that must be addressed. In the words of Ernest Boyer, "It is the scholarship of teaching that keeps the flame of scholarship alive" (Glassick, Huber, & Maeroff, 1997, p. 2).

APPENDIX A
Teaching Effectiveness Items.

Teaching Effectiveness Items.

- (1) Clearly stated the objectives of the course.
- (2) Seemed genuinely concerned with students' progress and was actively helpful when students were experiencing difficulties.
- (3) Told students how they would be evaluated.
- (4) Assigned grades fairly and impartially.
- (5) Accomplished the objectives of the course.
- (6) Appear to be well-prepared for class sessions.
- (7) Summarized or emphasized major points in lectures, discussion, or demonstrations.
- (8) Presented course material in an orderly manner.
- (9) Used examples or illustrations to clarify material.
- (10) Encouraged students to ask questions.
- (11) Was responsive to and clear in answering student questions.
- (12) Stimulated my thinking and interest.
- (13) The course has given me a better understanding of the subject.
- (14) How would you rate the overall performance of your instructor?

REFERENCES

REFERENCES

- Abrami, P. C. & D'Apollonia, S. (1990). The dimensionality of ratings and their use in personnel decisions. In M. Theall & Franklin, J.(Eds), Student ratings of instruction: Issues for improving practice (pp.97-111). San Francisco: Jossey-Bass.
- Abrami, P. C., D'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. Journal of Educational Psychology, 82, 219-231.
- Bolton, B., Bona, D. & Hinman, S. (1976). Dimensions of students' evaluations of instruction: An empirical synthesis. Psychological Reports, 58, 119-123.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 51, 281-309.
- Cohen, P. A. (1983). Comments on a selective review of the validity of student ratings of teaching. Journal of Higher Education, 54, 448-458.
- Cohen, P. A. (1986). An update and expanded meta-analysis of multisection student rating of validity studies. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Cohen, P. A. (1987). A critical analysis and reanalysis of the multisection validity meta-analysis. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.
- Doyle, K. O. & Whitely, S. E. (1974). Student ratings as criteria for effective teaching. American Educational Research Journal, 11, 259-274.
- Fernandez, J. & Mateo, M. A. (1992). Student evaluation of university teaching quality: Analysis of a questionnaire for a sample of university students in Spain. Educational and Psychological Measurement, 52, 675-686.
- El-Hassan, K. (1995). Students' ratings of instruction: Generalizability of findings. Studies in Educational Evaluation, 21, 411-429.

Frey, P. W. (1978). A two-dimensional analysis of student ratings of instruction. Research in Higher Education, 9, 69-91.

Glassick, C. E., Huber, M. T., & Maeroff (1997) *Scholarship Assessed: Evaluation of the Professoriate*. San Francisco: Jossey-Bass.

Gorsuch, R. L. (1983) *Factor Analysis*. London: Lawrence Erlbaum Associates.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. (1964). Dimensions of student evaluations of teaching. Journal of Educational Psychology, 55, 344-351.

Jeffreys, M. R., Massoni, M., O'Donnell, M. & Smolaka, I. (1997). Student evaluation of courses: Determining the reliability and validity of three survey instruments. Journal of Nursing Education, 36, 397-400.

Marsh, H. W. (1991). A multidimensional perspective on students' evaluations of teaching effectiveness: Reply to Abrami and d'Apollonia (1991). Journal of Educational Psychology, 83, 416-421.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

Marsh, H. W. (1984) Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, 76, 707-754.

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. Journal of Educational Psychology, 75, 150-166.

Marsh, H. W. & Hocevar, D. (1991). The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. Teaching & Teacher Education, 7, 9-18.

Novick, R. M. & Lewis C. (1967). Coefficient alpha and the reliability of composite measurements. Psychometrika, 32, 1-13

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.

Ory, J., & Parker, S. (1989). A survey of assessment activities at large research universities. Research in Higher Education, 30, 373-383.

Perry, R. P. (1990). Introduction to the special section. Journal of Educational Psychology, 82, 183-188.

Seldin, P. (Ed.). (1984). Changing practices in faculty evaluation: A critical assessment and recommendations for improvement. San Francisco: Jossey-Bass.

Statistical Package for Social Sciences (SPSS, Version 8.0). Chicago, IL.

Theall, M. & Franklin, J. (1990). Student Ratings in the context of complex evaluation systems. In M. Theall & J. Franklin (Eds), Student ratings of instruction: Issues for improving practice (pp.17-34). San Francisco: Jossey-Bass.